**Дисперсионный анализ**

Курсовая работа по дисциплине: «Системный анализ»

Исполнитель студент гр. 99 ИСЭ-2 Жбанов В.В.

Оренбургский государственный университет

Факультет информационных технологий

Кафедра прикладной информатики

г. Оренбург-2003

**Введение**

Цель работы: познакомится с таким статистическим методом, как дисперсионный анализ.

Дисперсионный анализ (от латинского Dispersio – рассеивание) – статистический метод, позволяющий анализировать влияние различных факторов на исследуемую переменную. Метод был разработан биологом Р. Фишером в 1925 году и применялся первоначально для оценки экспериментов в растениеводстве. В дальнейшем выяснилась общенаучная значимость дисперсионного анализа для экспериментов в психологии, педагогике, медицине и др.

Целью дисперсионного анализа является проверка значимости различия между средними с помощью сравнения дисперсий. Дисперсию измеряемого признака разлагают на независимые слагаемые, каждое из которых характеризует влияние того или иного фактора или их взаимодействия. Последующее сравнение таких слагаемых позволяет оценить значимость каждого изучаемого фактора, а также их комбинации /1/.

При истинности нулевой гипотезы (о равенстве средних в нескольких группах наблюдений, выбранных из генеральной совокупности), оценка дисперсии, связанной с внутригрупповой изменчивостью, должна быть близкой к оценке межгрупповой дисперсии.

При проведении исследования рынка часто встает вопрос о сопоставимости результатов. Например, проводя опросы по поводу потребления какого-либо товара в различных регионах страны, необходимо сделать выводы, на сколько данные опроса отличаются или не отличаются друг от друга. Сопоставлять отдельные показатели не имеет смысла и поэтому процедура сравнения и последующей оценки производится по некоторым усредненным значениям и отклонениям от этой усредненной оценки. Изучается вариация признака. За меру вариации может быть принята дисперсия. Дисперсия σ2 – мера вариации, определяемая как средняя из отклонений признака, возведенных в квадрат.

На практике часто возникают задачи более общего характера – задачи проверки существенности различий средних выборочных нескольких совокупностей. Например, требуется оценить влияние различного сырья на качество производимой продукции, решить задачу о влиянии количества удобрений на урожайность с/х продукции.

Иногда дисперсионный анализ применяется, чтобы установить однородность нескольких совокупностей (дисперсии этих совокупностей одинаковы по предположению; если дисперсионный анализ покажет, что и математические ожидания одинаковы, то в этом смысле совокупности однородны). Однородные же совокупности можно объединить в одну и тем самым получить о ней более полную информацию, следовательно, и более надежные выводы /2/.

**1 Дисперсионный анализ**

**1.1 Основные понятия дисперсионного анализа**

В процессе наблюдения за исследуемым объектом качественные факторы произвольно или заданным образом изменяются. Конкретная реализация фактора (например, определенный температурный режим, выбранное оборудование или материал) называется **уровнем фактора** или **способом обработки**. Модель дисперсионного анализа с фиксированными уровнями факторов называют **моделью I**, модель со случайными факторами - **моделью II**. Благодаря **варьированию**  фактора можно исследовать его влияние на величину отклика. В настоящее время общая теория дисперсионного анализа разработана для моделей I.

В зависимости от количества факторов, определяющих вариацию результативного признака, дисперсионный анализ подразделяют на однофакторный и многофакторный.

Основными схемами организации исходных данных с двумя и более факторами являются:

- **перекрестная классификация**, характерная для моделей I, в которых каждый уровень одного фактора сочетается при планировании эксперимента с каждой градацией другого фактора;

- **иерархическая (гнездовая) классификация**, характерная для модели II, в которой каждому случайному, наудачу выбранному значению одного фактора соответствует свое подмножество значений второго фактора.

Если одновременно исследуется зависимость отклика от качественных и количественных факторов, т.е. факторов **смешанной природы**, то используется **ковариационный анализ** /3/.

При обработке данных эксперимента наиболее разработанными и поэтому распространенными считаются две модели. Их различие обусловлено спецификой планирования самого эксперимента. В модели дисперсионного анализа с фиксированными эффектами исследователь намеренно устанавливает строго определенные уровни изучаемого фактора. Термин «фиксированный эффект» в данном контексте имеет тот смысл, что самим исследователем фиксируется количество уровней фактора и различия между ними. При повторении эксперимента он или другой исследователь выберет те же самые уровни фактора. В модели со случайными эффектами уровни значения фактора выбираются исследователем случайно из широкого диапазона значений фактора, и при повторных экспериментах, естественно, этот диапазон будет другим.

Таким образом, данные модели отличаются между собой способом выбора уровней фактора, что, очевидно, в первую очередь влияет на возможность обобщения полученных экспериментальных результатов. Для дисперсионного анализа однофакторных экспериментов различие этих двух моделей не столь существенно, однако в многофакторном дисперсионном анализе оно может оказаться весьма важным.

При проведении дисперсионного анализа должны выполняться следующие **статистические допущения:** независимо от уровня фактора величины отклика имеют нормальный (Гауссовский) закон распределения и одинаковую дисперсию. Такое равенство дисперсий называется гомогенностью. Таким образом, изменение способа обработки сказывается лишь на положении случайной величины отклика, которое характеризуется средним значением или медианой. Поэтому все наблюдения отклика принадлежат **сдвиговому семейству нормальных распределений**.

Говорят, что техника дисперсионного анализа является "робастной". Этот термин, используемый статистиками, означает, что данные допущения могут быть в некоторой степени нарушены, но несмотря на это, технику можно использовать.

При неизвестном законе распределения величин отклика используют непараметрические (чаще всего ранговые) методы анализа.

В основе дисперсионного анализа лежит разделение дисперсии на части или компоненты. Вариацию, обусловленную влиянием фактора, положенного в основу группировки, характеризует межгрупповая дисперсия σ2. Она является мерой вариации частных средних по группам вокруг общей средней и определяется по формуле:

,

где k - число групп;

 nj  - число единиц в j-ой группе;

 - частная средняя по j-ой группе;

 - общая средняя по совокупности единиц.

Вариацию, обусловленную влиянием прочих факторов, характеризует в каждой группе внутригрупповая дисперсия σj2.

.

Между общей дисперсией σ02, внутригрупповой дисперсией σ2 и межгрупповой дисперсией существует соотношение:

σ02 = + σ2.

Внутригрупповая дисперсия объясняет влияние неучтенных при группировке факторов, а межгрупповая дисперсия объясняет влияние факторов группировки на среднее значение по группе /2/.

**1.2 Однофакторный дисперсионный анализ**

Однофакторная дисперсионная модель имеет вид:

xij = μ + Fj + εij, (1)

где хij – значение исследуемой переменой, полученной на i-м уровне фактора (i=1,2,...,т) c j-м порядковым номером (j=1,2,...,n);

Fi – эффект, обусловленный влиянием i-го уровня фактора;

εij– случайная компонента, или возмущение, вызванное влиянием неконтролируемых факторов, т.е. вариацией переменой внутри отдельного уровня.

Основные предпосылки дисперсионного анализа:

- математическое ожидание возмущения εij равно нулю для любых i, т.е.

M(εij) = 0; (2)

- возмущения εij взаимно независимы;

- дисперсия переменной xij (или возмущения εij) постоянна для

любых i, j, т.е.

 D(εij) = σ2; (3)

- переменная xij (или возмущение εij) имеет нормальный закон

распределения N(0;σ2).

Влияние уровней фактора может быть как фиксированным или систематическим (модель I), так и случайным (модель II).

Пусть, например, необходимо выяснить, имеются ли существенные различия между партиями изделий по некоторому показателю качества, т.е. проверить влияние на качество одного фактора - партии изделий. Если включить в исследование все партии сырья, то влияние уровня такого фактора систематическое (модель I), а полученные выводы применимы только к тем отдельным партиям, которые привлекались при исследовании. Если же включить только отобранную случайно часть партий, то влияние фактора случайное (модель II). В многофакторных комплексах возможна смешанная модель III, в которой одни факторы имеют случайные уровни, а другие – фиксированные.

Пусть имеется m партий изделий. Из каждой партии отобрано соответственно n1, n2, …, nm изделий (для простоты полагается, что n1=n2=...=nm=n). Значения показателя качества этих изделий представлены в матрице наблюдений:

x11 x12 … x1n

x21 x22 … x2n

………………… = (xij), (i = 1,2, …, m; j = 1,2, …, n).

xm1 xm2 … xmn

Необходимо проверить существенность влияния партий изделий на их качество.

Если полагать, что элементы строк матрицы наблюдений – это численные значения случайных величин Х1,Х2,...,Хm, выражающих качество изделий и имеющих нормальный закон распределения с математическими ожиданиями соответственно a1,а2,...,аm и одинаковыми дисперсиями σ2, то данная задача сводится к проверке нулевой гипотезы Н0: a1=a2 =...= аm, осуществляемой в дисперсионном анализе.

Усреднение по какому-либо индексу обозначено звездочкой (или точкой) вместо индекса, тогда средний показатель качества изделий i-й партии, или групповая средняя для i-го уровня фактора, примет вид:

 , (4)

где i\* – среднее значение по столбцам;

ij – элемент матрицы наблюдений;

 n – объем выборки.

А общая средняя:

 . (5)

Сумма квадратов отклонений наблюдений хij от общей средней \*\* выглядит так:

2=2+2+

+22. (6)

или

Q = Q1 + Q2 + Q3.

Последнее слагаемое равно нулю

 =0. (7)

так как сумма отклонений значений переменной от ее средней равна нулю, т.е.

2=0.

Первое слагаемое можно записать в виде:

В результате получается тождество:

 Q = Q1 + Q2, (8)

где - общая, или полная, сумма квадратов отклонений;

- сумма квадратов отклонений групповых средних от общей средней, или межгрупповая (факторная) сумма квадратов отклонений;

- сумма квадратов отклонений наблюдений от групповых средних, или внутригрупповая (остаточная) сумма квадратов отклонений.

В разложении (8) заключена основная идея дисперсионного анализа. Применительно к рассматриваемой задаче равенство (8) показывает, что общая вариация показателя качества, измеренная суммой Q, складывается из двух компонент – Q1 и Q2, характеризующих изменчивость этого показателя между партиями (Q1) и изменчивость внутри партий (Q2), характеризующих одинаковую для всех партий вариацию под воздействием неучтенных факторов.

В дисперсионном анализе анализируются не сами суммы квадратов отклонений, а так называемые средние квадраты, являющиеся несмещенными оценками соответствующих дисперсий, которые получаются делением сумм квадратов отклонений на соответствующее число степеней свободы.

Число степеней свободы определяется как общее число наблюдений минус число связывающих их уравнений. Поэтому для среднего квадрата s12, являющегося несмещенной оценкой межгрупповой дисперсии, число степеней свободы k1=m-1, так как при его расчете используются m групповых средних, связанных между собой одним уравнением (5). А для среднего квадрата s22, являющегося несмещенной оценкой внутригрупповой дисперсии, число степеней свободы k2=mn-m, т.к. при ее расчете используются все mn наблюдений, связанных между собой m уравнениями (4).

Таким образом:

= Q1/(m-1),

= Q2/(mn-m).

Если найти математические ожидания средних квадратов и , подставить в их формулы выражение xij (1) через параметры модели, то получится:

 (9)

т.к. с учетом свойств математического ожидания

 а

 (10)

Для модели I с фиксированными уровнями фактора Fi(i=1,2,...,m) – величины неслучайные, поэтому

M(S) =2 /(m-1) +σ2.

Гипотеза H0 примет вид Fi = F\*(i = 1,2,...,m), т.е. влияние всех уровней фактора одно и то же. В случае справедливости этой гипотезы

M(S)= M(S)= σ2.

Для случайной модели II слагаемое Fi в выражении (1) – величина случайная. Обозначая ее дисперсией

получим из (9)

 (11)

и, как и в модели I

M(S)= σ2.

В таблице 1.1 представлен общий вид вычисления значений, с помощью дисперсионного анализа.

Таблица 1.1 – Базовая таблица дисперсионного анализа

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Компоненты дисперсии | Сумма квадратов | Число степеней свободы | Средний квадрат | Математическое ожидание среднего квадрата |
| Межгрупповая |  | m-1 | = Q1/(m-1) |  |
| Внутригрупповая |  | mn-m | = Q2/(mn-m) | M(S)= σ2 |
| Общая |  | mn-1 |  |  |

Гипотеза H0 примет вид σF2 =0. В случае справедливости этой гипотезы

M(S)= M(S)= σ2.

В случае однофакторного комплекса как для модели I, так и модели II средние квадраты S2 и S2, являются несмещенными и независимыми оценками одной и той же дисперсии σ2.

Следовательно, проверка нулевой гипотезы H0 свелась к проверке существенности различия несмещенных выборочных оценок S и S дисперсии σ2.

Гипотеза H0 отвергается, если фактически вычисленное значение статистики F = S/S больше критического Fα:K1:K2, определенного на уровне значимости α при числе степеней свободы k1=m-1 и k2=mn-m, и принимается, если F < Fα:K1:K2 .

F- распределение Фишера (для x > 0) имеет следующую функцию плотности (для = 1, 2, ...; = 1, 2, ...):

где - степени свободы;

Г - гамма-функция.

Применительно к данной задаче опровержение гипотезы H0 означает наличие существенных различий в качестве изделий различных партий на рассматриваемом уровне значимости.

Для вычисления сумм квадратов Q1, Q2, Q часто бывает удобно использовать следующие формулы:

 (12)

 (13)

 (14)

т.е. сами средние, вообще говоря, находить не обязательно.

 Таким образом, процедура однофакторного дисперсионного анализа состоит в проверке гипотезы H0 о том, что имеется одна группа однородных экспериментальных данных против альтернативы о том, что таких групп больше, чем одна. Под однородностью понимается одинаковость средних значений и дисперсий в любом подмножестве данных. При этом дисперсии могут быть как известны, так и неизвестны заранее. Если имеются основания полагать, что известная или неизвестная дисперсия измерений одинакова по всей совокупности данных, то задача однофакторного дисперсионного анализа сводится к исследованию значимости различия средних в группах данных /1/.

**1.3 Многофакторный дисперсионный анализ**

Следует сразу же отметить, что принципиальной разницы между многофакторным и однофакторным дисперсионным анализом нет. Многофакторный анализ не меняет общую логику дисперсионного анализа, а лишь несколько усложняет ее, поскольку, кроме учета влияния на зависимую переменную каждого из факторов по отдельности, следует оценивать и их совместное действие. Таким образом, то новое, что вносит в анализ данных многофакторный дисперсионный анализ, касается в основном возможности оценить межфакторное взаимодействие. Тем не менее, по-прежнему остается возможность оценивать влияние каждого фактора в отдельности. В этом смысле процедура многофакторного дисперсионного анализа (в варианте ее компьютерного использования) несомненно более экономична, поскольку всего за один запуск решает сразу две задачи: оценивается влияние каждого из факторов и их взаимодействие /3/.

Общая схема двухфакторного эксперимента, данные которого обрабатываются дисперсионным анализом имеет вид:

Взаимодействие факторов A и B

Фактор B:

3 уровня

Фактор А:

2 уровня

Прочие неучитываемые (случайные) факторы

Рисунок 1.1 – Схема двухфакторного эксперимента

Данные, подвергаемые многофакторному дисперсионному анализу, часто обозначают в соответствии с количеством факторов и их уровней.

Предположив, что в рассматриваемой задаче о качестве различных m партий изделия изготавливались на разных t станках и требуется выяснить, имеются ли существенные различия в качестве изделий по каждому фактору:

А - партия изделий;

B - станок.

В результате получается переход к задаче двухфакторного дисперсионного анализа.

Все данные представлены в таблице 1.2, в которой по строкам - уровни Ai фактора А, по столбцам — уровни Bj фактора В, а в соответствующих ячейках, таблицы находятся значения показателя качества изделий xijk (i=1,2,...,m; j=1,2,...,l; k=1,2,...,n).

Таблица 1.2 – Показатели качества изделий

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | B1 | B2 | … | Bj | … | Bl |
| A1 | x11l,…,x11k | x12l,…,x12k | … | x1jl,…,x1jk | … | x1ll,…,x1lk |
| A2 | x21l,…,x21k | x22l,…,x22k | … | x2jl,…,x2jk | … | x2ll,…,x2lk |
| … | … | … | … | … | … | … |
| Ai | xi1l,…,xi1k | xi2l,…,xi2k | … | xijl,…,xijk | … | xjll,…,xjlk |
| … | … | … | … | … | … | … |
| Am | xm1l,…,xm1k | xm2l,…,xm2k | … | xmjl,…,xmjk | … | xmll,…,xmlk |

Двухфакторная дисперсионная модель имеет вид:

 xijk=μ+Fi+Gj+Iij+εijk, (15)

где xijk - значение наблюдения в ячейке ij с номером k;

μ - общая средняя;

Fi - эффект, обусловленный влиянием i-го уровня фактора А;

Gj - эффект, обусловленный влиянием j-го уровня фактора В;

Iij - эффект, обусловленный взаимодействием двух факторов, т.е. отклонение от средней по наблюдениям в ячейке ij от суммы первых трех слагаемых в модели (15);

εijk - возмущение, обусловленное вариацией переменной внутри отдельной ячейки.

Предполагается, что εijk имеет нормальный закон распределения N(0; с2), а все математические ожидания F\*,G\*, Ii\*, I\*j равны нулю.

Групповые средние находятся по формулам:

- в ячейке:

,

по строке:

по столбцу:

общая средняя:

В таблице 1.3 представлен общий вид вычисления значений, с помощью дисперсионного анализа.

Таблица 1.3 – Базовая таблица дисперсионного анализа

|  |  |  |  |
| --- | --- | --- | --- |
| Компоненты дисперсии | Сумма квадратов | Число степеней свободы | Средние квадраты |
| Межгрупповая (фактор А) |  | m-1 |  |
| Межгрупповая (фактор B) |  | l-1 |  |
| Взаимодействие |  | (m-1)(l-1) |  |
| Остаточная |  | mln - ml |  |
| Общая |  | mln - 1 |  |

Проверка нулевых гипотез HA, HB, HAB об отсутствии влияния на рассматриваемую переменную факторов А, B и их взаимодействия AB осуществляется сравнением отношений , , (для модели I с фиксированными уровнями факторов) или отношений , , (для случайной модели II) с соответствующими табличными значениями F – критерия Фишера – Снедекора. Для смешанной модели III проверка гипотез относительно факторов с фиксированными уровнями производится также как и в модели II, а факторов со случайными уровнями – как в модели I.

Если n=1, т.е. при одном наблюдении в ячейке, то не все нулевые гипотезы могут быть проверены так как выпадает компонента Q3 из общей суммы квадратов отклонений, а с ней и средний квадрат , так как в этом случае не может быть речи о взаимодействии факторов.

С точки зрения техники вычислений для нахождения сумм квадратов Q1, Q2, Q3, Q4, Q целесообразнее использовать формулы:

Q3 = Q – Q1 – Q2 – Q4.

Отклонение от основных предпосылок дисперсионного анализа — нормальности распределения исследуемой переменной и равенства дисперсий в ячейках (если оно не чрезмерное) — не сказывается существенно на результатах дисперсионного анализа при равном числе наблюдений в ячейках, но может быть очень чувствительно при неравном их числе. Кроме того, при неравном числе наблюдений в ячейках резко возрастает сложность аппарата дисперсионного анализа. Поэтому рекомендуется планировать схему с равным числом наблюдений в ячейках, а если встречаются недостающие данные, то возмещать их средними значениями других наблюдений в ячейках. При этом, однако, искусственно введенные недостающие данные не следует учитывать при подсчете числа степеней свободы /1/.

**2 Применение дисперсионного анализа в различных процессах и исследованиях**

**2.1 Использование дисперсионного анализа при изучении миграционных процессов**

Миграция - сложное социальное явление, во многом определяющее экономическую и политическую стороны жизни общества. Исследование миграционных процессов связано с выявлением факторов заинтересованности, удовлетворенности условиями труда, и оценкой влияния полученных факторов на межгрупповое движение населения.

λij=ciqijaj,

где λij – интенсивность переходов из исходной группы i (выхода) в новую j (входа);

ci – возможность и способности покинуть группу i (ci≥0);

qij – привлекательность новой группы по сравнению с исходной (0≤qij≤1);

aj – доступность группы j (aj≥0).

Если считать численность группы i равной ni, то оценкой случайной величины νij - числа переходов из i в j – будет niciqijaj:

 νij≈ niλij=niciqijaj. (16)

На практике для отдельного человека вероятность p перехода в другую группу мала, а численность рассматриваемой группы n велика. В этом случае действует закон редких событий, то есть пределом νij является распределение Пуассона с параметром μ=np:

.

С ростом μ распределение приближается к нормальному. Преобразованную же величину √νij можно считать нормально распределенной.

Если прологарифмировать выражение (16) и сделать необходимые замены переменных, то можно получить модель дисперсионного анализа:

ln√νij=½lnνij=½(lnni+lnci+lnqij+lnaj)+εij,

Xi,j=2ln√νij-lnni-lnqij,

Ci=lnci,

Aj=lnaj,

Xi,j=Ci+Aj+ε.

Значения Ci и Aj позволяют получить модель двухфакторного дисперсионного анализа с одним наблюдением в клетке. Обратным преобразованием из Ci и Aj вычисляются коэффициенты ci и aj.

При проведении дисперсионного анализа в качестве значений результативного признака Y следует взять величины:

Yij=Xi,j-X,

Х=(Х1,1+Х1,2+:+Хmi,mj)/mimj,

где mimj- оценка математического ожидания Хi,j;

 Хmi и Хmj - соответственно количество групп выхода и входа.

Уровнями фактора I будут mi групп выхода, уровнями фактора J - mj групп входа. Предполагается mi=mj=m. Встает задача проверки гипотез HI и HJ о равенствах математических ожиданий величины Y при уровнях Ii и при уровнях Jj, i,j=1,…,m. Проверка гипотезы HI основывается на сравнении величин несмещенных оценок дисперсии sI2 и so2. Если гипотеза HI верна, то величина F(I)= sI 2/so2 имеет распределение Фишера с числами степеней свободы k1=m-1 и k2=(m-1)(m-1). Для заданного уровня значимости α находится правосторонняя критическая точка xпр,αкр. Если числовое значение F(I)чис величины попадает в интервал(xпр,αкр, +∞), то гипотеза HI отвергается и считается, что фактор I влияет на результативный признак. Степень этого влияния по результатам наблюдений измеряется выборочным коэффициентом детерминации, который показывает, какая доля дисперсии результативного признака в выборке обусловлена влиянием на него фактора I. Если же F(I)чис<xпр,αкр, то гипотеза HI не отвергаются и считаются, что влияние фактора I не подтвердилось. Аналогично проверяется гипотеза HJ о влиянии фактора J /4/.

**2.2 Принципы математико-статистического анализа данных медико-биологических исследований**

В зависимости от поставленной задачи, объема и характера материала, вида данных и их связей находится выбор методов математической обработки на этапах как предварительного (для оценки характера распределения в исследуемой выборке), так и окончательного анализа в соответствии с целями исследования. Крайне важным аспектом является проверка однородности выбранных групп наблюдения, в том числе контрольных, что может быть проведено или экспертным путем, или методами многомерной статистики (например, с помощью кластерного анализа). Но первым этапом является составление вопросника, в котором предусматривается стандартизованное описание признаков. В особенности при проведении эпидемиологических исследований, где необходимо единство в понимании и описании одних и тех же симптомов разными врачами, включая учет диапазонов их изменений (степени выраженности). В случае существенности различий в регистрации исходных данных (субъективная оценка характера патологических проявлений различными специалистами) и невозможности их приведения к единому виду на этапе сбора информации, может быть затем осуществлена так называемая коррекция ковариант, которая предполагает нормализацию переменных, т.е. устранение ненормальностей показателей в матрице данных. "Согласование мнений" осуществляется с учетом специальности и опыта врачей, что позволяет затем сравнивать полученные ими результаты обследования между собой. Для этого могут использоваться многомерный дисперсионный и регрессионный анализы.

Признаки могут быть как однотипными, что бывает редко, так и разнотипными. Под этим термином понимается их различная метрологическая оценка. Количественные или числовые признаки - это замеренные в определенной шкале и в шкалах интервалов и отношений (I группа признаков). Качественные, ранговые или балльные используются для выражения медицинских терминов и понятий не имеющих цифровых значений (например, тяжесть состояния) и замеряются в шкале порядка (II группа признаков). Классификационные или номинальные (например, профессия, группа крови) - это замеренные в шкале наименований (III группа признаков).

Во многих случаях делается попытка анализа крайне большого числа признаков, что должно способствовать повышению информативности представленной выборки. Однако выбор полезной информации, то есть осуществление отбора признаков является операцией совершенно необходимой, поскольку для решения любой классификационной задачи должны быть отобраны сведения, несущие полезную для данной задачи информацию. В случае, если это не осуществлено по каким-то причинам исследователем самостоятельно или отсутствуют достаточно обоснованные критерии для снижения размерности пространства признаков по содержательным соображениям, борьба с избыточностью информации осуществляется уже формальными методами путем оценки информативности.

Дисперсионный анализ позволяет определить влияние разных факторов (условий) на исследуемый признак (явление), что достигается путем разложения совокупной изменчивости (дисперсии, выраженной в сумме квадратов отклонений от общего среднего) на отдельные компоненты, вызванные влиянием различных источников изменчивости.

С помощью дисперсионного анализа исследуются угрозы заболевания при наличии факторов риска. Концепция относительного риска рассматривает отношение между пациентами с определенной болезнью и не имеющими ее. Величина относительного риска дает возможность определить, во сколько раз увеличивается вероятность заболеть при его наличии, что может быть оценено с помощью следующей упрощенной формулы:

r' = a\*d / b\*c,

где a - наличие признака в исследуемой группе;

b - отсутствие признака в исследуемой группе;

c - наличие признака в группе сравнения (контрольной);

d - отсутствие признака в группе сравнения (контрольной).

Показатель атрибутивного риска (rA) служит для оценки доли заболеваемости, связанной с данным фактором риска:

,

где Q - частота признака, маркирующего риск, в популяции;

 r' - относительный риск.

Выявление факторов, способствующих возникновению (проявлению) заболевания, т.е. факторов риска может осуществляться различными способами, например, путем оценки информативности с последующим ранжированием признаков, что однако не указывает на совокупное действие отобранных параметров, в отличие от применения регрессионного, факторного анализов, методов теории распознавания образов, которые дают возможность получать "симптомокомплексы" риск-факторов. Кроме того, более сложные методы позволяют анализировать и непрямые связи между факторами риска и заболеваниями /5/.

**2.3 Биотестирование почвы**

Многообразные загрязняющие вещества, попадая в агроценоз, могут претерпевать в нем различные превращения, усиливая при этом свое токсическое действие. По этой причине оказались необходимыми методы интегральной оценки качества компонентов агроценоза. Исследования проводили на базе многофакторного дисперсионного анализа в 11-ти польном зернотравянопропашном севообороте. В опыте изучалось влияние следующих факторов: плодородие почвы (А), система удобрений (В), система защиты растений (С). Плодородие почвы, система удобрений и система защиты растений изучались в дозах 0, 1, 2 и 3. Базовые варианты были представлены следующими комбинациями:

000 - исходный уровень плодородия, без применения удобрений и средств защиты растений от вредителей , болезней и сорняков;

111 - средний уровень плодородия почвы, минимальная доза удобрения, биологическая защита растений от вредителей и болезней;

222 - исходный уровень плодородия почвы, средняя доза удобрений, химическая защита растений от сорняков;

333 - высокий уровень плодородия почвы, высокая доза удобрений, химическая защита растений от вредителей и болезней.

Изучались варианты, где представлен только один фактор:

200 – плодородие:

020 – удобрения;

002 - средства защиты растений.

А также варианты с различным сочетанием факторов - 111, 131, 133, 022, 220, 202, 331, 313, 311.

Целью исследования являлось изучение торможения хлоропластов и коэффициента мгновенного роста, как показателей загрязнения почвы, в различных вариантах многофакторного опыта.

Торможение фототаксиса хлоропластов ряски малой исследовали в различных горизонтах почвы: 0-20, 20-40 см. Анализ изменчивости фототаксиса в разных вариантах опыта показал достоверное влияние каждого из факторов (плодородия почвы, системы удобрений и системы защиты растений). Доля в общей дисперсии плодородия почвы составила 39,7%, системы удобрений - 30,7%, системы защиты растений - 30,7 %.

Для исследования совокупного влияния факторов на торможение фототаксиса хлоропластов использовались различные сочетания вариантов опыта: в первом случае - 000, 002, 022, 222, 220, 200, 202, 020, во втором случае - 111, 333, 331, 313, 133, 311, 131.

Результаты двухфакторного дисперсионного анализа свидетельствуют о достоверном влиянии взаимодействующих системы удобрений и системы защиты растений на различия в фототаксисе для первого случая (доля в общей дисперсии составила 10,3%). Для второго случая обнаружено достоверное влияние взаимодействующих плодородия почвы и системы удобрений (53,2%).

Трехфакторный дисперсионный анализ показал в первом случае достоверное влияние взаимодействия всех трех факторов. Доля в общей дисперсии составила 47,9%.

Коэффициент мгновенного роста исследовали в различных вариантах опыта 000, 111, 222, 333, 002, 200, 220. Первый этап тестирования - до внесения гербицидов на посевах озимой пшеницы (апрель), второй этап - после внесения гербицидов (май) и последний - на момент уборки (июль). Предшетвенники - подсолнечник и кукуруза на зерно.

Появление новых листецов наблюдали после короткой лаг-фазы с периодом суммарного удвоения сырой массы 2 - 4 суток.

В контроле и в каждом варианте на основании полученных результатов рассчитывали коэффициент мгновенного роста популяции r и далее рассчитывали время удвоения численности листецов (t удв ).

tудв=ln2/r.

Расчет этих показателей был проведен в динамике с анализом почвенных образцов. Анализ данных показал, что время удвоения популяции рясок до обработки почвы было наименьшем по сравнению с данными после обработки и на момент уборки. В динамике наблюдений больший интерес вызывает отклик почвы после внесения гербицида и на момент уборки. Прежде всего взаимодействие с удобрениями и уровнем плодородия.

Подчас получить прямой отклик на внесение химических препараратов может быть осложнено взаимодействием препарата с удобрениями, как органическими, так и минеральными. Полученные данные позволили проследить динамику отклика вносимых препаратов, во всех вариантах с химическими средствами защиты, где отмечается приостановка роста индикатора.

Данные однофакторного дисперсионного анализа показали достоверное влияние каждого показателя на темпы роста ряски малой на первом этапе. На втором этапе эффект различий по плодородию почвы составил 65,0 %, по системе удобрений и системе защиты растений - по 65,0%. Факторы показали достоверные различия среднего по коэффициенту мгновенного роста варианта 222 и вариантов 000, 111, 333. На третьем этапе доля в общей дисперсии плодородия почвы составила 42,9%, системы удобрений и системы защиты растений - по 42,9%. Отмечено достоверное различие по средним значениям вариантов 000 и 111, вариантов 333 и 222.

Исследуемые образцы почвы с вариантов полевого мониторинга отличаются друг от друга по показателю торможение фототаксиса. Отмечено влияние факторов плодородия, система удобрений и средства защиты растений с долями 30,7 и 39,7% при однофакторном анализе, при двух факторном и трехфакторном - зарегистрировали совместное влияние факторов.

Анализ результатов опыта показал незначительные различия между горизонтами почвы по показателю - торможение фототаксиса. Отличия отмечены по средним значениям.

На всех вариантах, где имеются средства защиты растений наблюдается изменения положения хлоропластов и приостановка роста ряски малой /6/.

**2.4 Грипп вызывает повышенную выработку гистамина**

Исследователи из детской больницы в Питсбурге (США) получили первые доказательства того, что при острых респираторных вирусных инфекциях повышается уровень гистамина. Несмотря на то, что и раньше предполагалось, что гистамин играет определенную роль в возникновении симптомов острых респираторных инфекциях верхних дыхательных путей.

 Ученых интересовало, почему многие люди применяют для самолечения «простудных» заболеваний и насморка антигистаминные препараты, которые во многих странах входят в категорию OTC, т.е. доступны без рецепта врача.

Целью проведенного исследования было определить, повышается ли продукция гистамина при экспериментальной инфекции, вызванной вирусом гриппа А.

15 здоровым добровольцам интраназально ввели вирус гриппа А, а затем наблюдали за развитием инфекции. Ежедневно в течение заболевания у добровольцев собиралась утренняя порция мочи, а затем проводилось определение гистамина и его метаболитов и рассчитывалось общее количество гистамина и его метаболитов, выделенных за сутки.

Заболевание развилось у всех 15 добровольцев. Дисперсионный анализ подтвердил достоверно более высокий уровень гистамина в моче на 2-5 сутки вирусной инфекции (p<0,02) - период, когда симптомы «простуды» наиболее выражены. Парный анализ показал, что наиболее значительно уровень гистамина повышается на 2 день заболевания. Кроме этого, оказалось, что суточное количество гистамина и его метаболитов в моче при гриппе примерно такое же, как и при обострении аллергического заболевания.

Результаты данного исследования служат первыми прямыми доказательствами того, что уровень гистамина повышается при острых респираторных инфекциях /7/.

**Дисперсионный анализ в химии**

Дисперсионный анализ – совокупность методов определения дисперсности, т. е. характеристики размеров частиц в дисперсных системах. Дисперсионный анализ включает различные способы определения размеров свободных частиц в жидких и газовых средах, размеров каналов-пор в тонкопористых телах (в этом случае вместо понятия дисперсности используют равнозначное понятие пористости), а также удельной поверхности. Одни из методов дисперсионного анализа позволяют получать полную картину распределения частиц по размерам (объёмам), а другие дают лишь усреднённую характеристику дисперсности (пористости).

К первой группе относятся, например, методы определения размеров отдельных частиц непосредственным измерением (ситовой анализ, оптическая и электронная микроскопия) или по косвенным данным: скорости оседания частиц в вязкой среде (седиментационный анализ в гравитационном поле и в центрифугах), величине импульсов электрического тока, возникающих при прохождении частиц через отверстие в непроводящей перегородке (кондуктометрический метод).

Вторая группа методов объединяет оценку средних размеров свободных частиц и определение удельной поверхности порошков и пористых тел. Средний размер частиц находят по интенсивности рассеянного света (нефелометрия), с помощью ультрамикроскопа, методами диффузии и т.д., удельную поверхность - по адсорбции газов (паров) или растворённых веществ, по газопроницаемости, скорости растворения и др. способами. Ниже приведены границы применимости различных методов дисперсионного анализа (размеры частиц в метрах):

Ситовой анализ – 10-2-10-4

Седиментационный анализ в гравитационном поле – 10-4-10-6

Кондуктометрический метод – 10-4-10-6

Микроскопия – 10-4-10-7

Метод фильтрации – 10-5-10-7

Центрифугирование – 10-6-10-8

Ультрацентрифугирование – 10-7-10-9

Ультрамикроскопия – 10-7-10-9

Нефелометрия – 10-7-10-9

Электронная микроскопия – 10-7-10-9

Метод диффузии – 10-7-10-10

Дисперсионный анализ широко используют в различных областях науки и промышленного производства для оценки дисперсности систем (суспензий, эмульсий, золей, порошков, адсорбентов и т.д.) с величиной частиц от нескольких миллиметров (10-3 м) до нескольких нанометров (10-9 м) /8/.

**2.6 Использование прямого преднамеренного внушения в бодрствующем состоянии в методике воспитания физических качеств**

Физическая подготовка – основополагающая сторона спортивной тренировки, так как в большей мере, чем другие стороны подготовки, характеризуется физическими нагрузками, воздействующими на морфофункциональные свойства организма. От уровня физической подготовленности зависят успешность технической подготовки, содержание тактики спортсмена, реализация личностных свойств в процессе тренировок и состязаний.

Одной из основных задач физической подготовки является воспитание физических качеств. В связи с этим возникает необходимость в разработке педагогических средств и методов, позволяющих учитывать возрастные особенности юных спортсменов, сохраняющих их здоровье, не требующих дополнительных затрат времени и в то же время стимулирующих рост физических качеств и, как следствие, - спортивного мастерства. Использование вербального гетеровоздействия в тренировочном процессе в группах начальной подготовки - одно из перспективных направлений исследований по данной проблеме.

Анализ теории и практики реализации внушающего вербального гетеровоздействия выявил основные противоречия:

- доказанность эффективного использования специфических методов вербального гетеровоздействия в тренировочном процессе и практическую невозможность их использования тренером;

- признание прямого преднамеренного внушения (далее ППВ) в бодрствующем состоянии как одного из основных методов вербального гетеровоздействия в педагогической деятельности тренера и отсутствие теоретического обоснования методических особенностей его применения в спортивной подготовке, и в частности в процессе воспитания физических качеств.

В связи с выявленными противоречиями и недостаточной разработанностью проблема использования системы методов вербального гетеровоздействия в процессе воспитания физических качеств спортсменов предопределила цель исследования - разработать рациональные целенаправленные методики ППВ в бодрствующем состоянии, способствующие совершенствованию процесса воспитания физических качеств на основе оценки психического состояния, проявления и динамики физических качеств дзюдоистов групп начальной подготовки.

С целью апробации и определения эффективности экспериментальных методик ППВ при воспитании физических качеств дзюдоистов был проведен сравнительный педагогический эксперимент, в котором приняли участие четыре группы – три экспериментальных и одна контрольная. В первой экспериментальной группе (ЭГ) использовалась методика ППВ М1, во второй - методика ППВ М2, в третьей - методика ППВ М3. В контрольной группе (КГ) методики ППВ не применялись.

Для определения эффективности педагогического воздействия методик ППВ в процессе воспитания у дзюдоистов физических качеств был проведен однофакторный дисперсионный анализ.

Степень влияния методики ППВ M1 в процессе воспитания:

- выносливости:

а) после третьего месяца составила 11,1%;

- скоростных способностей:

а) после первого месяца - 16,4%;

б) после второго - 26,5%;

в) после третьего - 34,8%;

- силы:

а) после второго месяца - 26, 7%;

б) после третьего - 35,3%;

- гибкости:

а) после третьего месяца - 20,8%;

- координационных способностей:

а) после второго месяца основного педагогического эксперимента степень влияния методики составила 6,4%;

б) после третьего - 10,2%.

Следовательно, существенные изменения в показателях уровня развития физических качеств с использованием методики ППВ М1 обнаружены в скоростных способностях и силе, степень влияния методики в данном случае наибольшая. Наименьшая степень влияния методики обнаружена в процессе воспитания выносливости, гибкости, координационных способностей, что дает основание говорить о недостаточной эффективности использования методики ППВ М1 при воспитании указанных качеств.

Степень влияния методики ППВ M2 в процессе воспитания:

- выносливости

а) после первого месяца эксперимента - 12,6%;

б) после второго - 17,8%;

в) после третьего - 20,3%.

- скоростных способностей:

а) после третьего месяца тренировочных занятий - 28%.

- силы:

а) после второго месяца - 27,9%;

б) после третьего - 35,9%.

- гибкости:

а) после третьего месяца тренировочных занятий - 14,9%;

- координационных способностей - 13,1%.

Полученный результат однофакторного дисперсионного анализа данной ЭГ позволяет сделать вывод о том, что методика ППВ М2 наиболее результативна при воспитании выносливости и силы. Менее эффективна она в процессе воспитания гибкости, скоростных и координационных способностей.

Степень влияния методики ППВ М3 в процессе воспитания:

- выносливости:

а) после первого месяца эксперимента 16,8%;

б) после второго - 29,5%;

в) после третьего - 37,6%.

- скоростных способностей:

а) после первого месяца - 26,3%;

б) после второго - 31,3%;

в) после третьего - 40,9%.

- силы:

а) после первого месяца - 18,7%;

б) после второго - 26,7%;

в) после третьего - 32,3%.

- гибкости:

а) после первого - изменений нет;

б) после второго - 16,9%;

в) после третьего - 23,5%.

- координационных способностей:

а) после первого месяца изменений нет;

б) после второго - 23,8%;

в) после третьего - 91% .

Таким образом, однофакторный дисперсионный анализ показал, что использование методики ППВ М3 в подготовительном периоде наиболее эффективно в процессе воспитания физических качеств, так как наблюдается увеличение степени ее влияния после каждого месяца педагогического эксперимента /9/.

**2.7 Купирование острой психотической симптоматики у больных шизофренией атипичным нейролептиком**

Цель исследования сводилась к изучению возможности применения рисполепта для купирования острых психозов у больных с диагнозом шизофрении (параноидный тип по МКБ-10) и шизоаффективного расстройства. При этом в качестве основного изучаемого критерия использовался показатель длительности сохранения психотической симптоматики в условиях фармакотерапии рисполептом (основная группа) и классическими нейролептиками.

Основные задачи исследования сводились к определению показателя длительности психоза (так называемый нетто-психоз), под которым понималось сохранение продуктивной психотической симптоматики с момента начала применения нейролептиков, выраженное в днях. Данный показатель был рассчитан отдельно для группы, принимавшей рисперидон, и отдельно для группы, принимавшей классические нейролептики.

Наряду с этим была поставлена задача по определению доли редукции продуктивной симптоматики под влиянием рисперидона в сравнении с классическими нейролептиками в разные сроки терапии.

В общей сложности изучены 89 больных (42 мужчины и 47 женщин) с острой психотической симптоматикой в рамках параноидной формы шизофрении (49 больных) и шизоаффективного расстройства (40 больных).

Первый эпизод и длительность заболевания до 1 года были зарегистрированы у 43 больных, тогда как в остальных случаях на момент исследования отмечались последующие эпизоды шизофрении при длительности заболевания свыше 1 года.

Терапию рисполептом получали 29 человек, среди которых с так называемым первым эпизодом было 15 больных. Терапию классическими нейролептиками получали 60 человек, среди которых с первым эпизодом было 28 человек. Доза рисполепта варьировала в диапазоне от 1 до 6 мг в сутки и в среднем составляла 4±0,4 мг/сут. Рисперидон принимали исключительно внутрь после еды один раз в сутки в вечернее время.

Терапия классическими нейролептиками включала применение трифлуоперазина (трифтазина) в суточной дозе до 30 мг внутримышечно, галоперидола в суточной дозе до 20 мг внутримышечно, триперидола в суточной дозе до 10 мг внутрь. Подавляющее большинство больных принимало классические нейролептики в виде монотерапии в течение первых двух недель, после чего переходили в случае необходимости (при сохранении бредовой, галлюцинаторной или другой продуктивной симптоматики) к сочетанию нескольких классических нейролептиков. При этом в качестве основного препарата оставался нейролептик с выраженным элективным антибредовым и антигаллюцинаторным аффектом (например, галоперидол или трифтазин), к нему присоединяли в вечернее время препарат с отчетливым гипноседативным эффектом (аминазин, тизерцин, хлорпротиксен в дозах до 50-100 мг/сут).

В группе, принимавшей классические нейролептики, был предусмотрен прием корректоров холинолитического ряда (паркопан, циклодол) в дозах до 10-12 мг/сут. Корректоры назначались в случае появления отчетливых побочных экстрапирамидных эффектов в виде острых дистоний, лекарственного паркинсонизма и акатизии.

В таблице 2.1 представлены данные по длительности психоза при лечении рисполептом и классическими нейролептиками.

Таблица 2.1 – Длительность психоза ("нетто-психоз") при лечении рисполептом и классическими нейролептиками

|  |  |  |
| --- | --- | --- |
| Группа больных | Длительность терапии, дни | Значимость различий, p |
| рисполепт | классические нейролептики |
| Вся выборка(n = 89)  | 15,4±1,4(n = 29) | 31,4±2,5(n = 60) | <0,001 |
| Больные с 1-м эпизодом(n = 43) | 14,7±1,9(n = 15) | 25,8±2,9(n = 28) | <0,002 |
| Больные с последующими эпизодами(n = 46) | 16,3±2,0(n = 14) | 35,1±3,8(n = 32) | <0,001 |
| Больные с галлюцинаторно-параноидным синдромом(n = 49) | 15,2±1,5(n = 23) | 34,7±4,4(n = 26) | <0,001 |
| Больные с шизоаффективным синдромом(n = 40) | 16,5±3,8(n = 6) | 28,9±2,9(n = 34) | <0,01 |

Как следует из данных таблицы, при сравнении длительности психоза при терапии классическими нейролептиками и рисперидоном наблюдается практически двукратное сокращение продолжительности психотической симптоматики под влиянием рисполепта. Существенно, что на данную величину продолжительности психоза не влияли ни факторы порядкового номера приступов, ни характер картины ведущего синдрома. Иначе говоря, длительность психоза определялась исключительно фактором терапии, т.е. зависела от типа применяемого препарата безотносительно порядкового номера приступа, продолжительности заболевания и характера ведущего психопатологического синдрома.

С целью подтверждения полученных закономерностей был проведен двухфакторный дисперсионный анализ. При этом поочередно учитывалось взаимодействие фактора терапии и порядкового номера приступа (1-й этап) и взаимодействие фактора терапии и характера ведущего синдрома (2-й этап). Результаты дисперсионного анализа подтвердили влияние фактора терапии на величину длительности психоза (F=18,8) при отсутствии влияния фактора номера приступа (F=2,5) и фактора типа психопатологического синдрома (F=1,7). Немаловажно, что совместное влияние фактора терапии и номера приступа на величину длительности психоза также отсутствовало, равно как и совместное влияние фактора терапии и фактора психопатологического синдрома.

Таким образом, результаты дисперсионного анализа подтвердили влияние только фактора применяемого нейролептика. Рисполепт однозначно приводил к сокращению длительности психотической симптоматики по сравнению с традиционными нейролептиками примерно в 2 раза. Принципиально, что этот эффект был достигнут, несмотря на пероральный прием рисполепта, тогда как классические нейролептики применялись у большей части больных парентерально /10/.

**2.8 Снование фасонной пряжи с ровничным эффектом**

В Костромском Государственном технологическом университете разработана новая структура фасонной нити с переменными геометрическими параметрами. В связи с этим возникает проблема переработки фасонной пряжи в приготовительном производстве. Данное исследование посвящалось процессу снования по вопросам: выбор типа натяжного устройства, дающего минимальный разброс натяжения и выравнивание натяжения, нитей различной линейной плотности по ширине сновального вала.

Объект исследования – льняная фасонная нить четырех вариантов линейной плотности от 140 до 205 текса. Исследовалась работа натяжных приборов трех типов: фарфорового шайбового, двухзонного НС-1П и однозонного НС-1П. Экспериментальное исследование натяжения снующихся нитей производилось на сновальной машине СП-140-3Л. Скорость снования, масса тормозных шайб соответствовали технологическим параметрам снования пряжи.

Для исследования зависимости натяжения фасонной нити от геометрических параметров при сновании проведен анализ для двух факторов: X1 - диаметр эффекта, X2 - длина эффекта. Выходными параметрами являются натяжение Y1 и колебание натяжения Y2.

Полученные уравнения регрессии адекватны экспериментальным данным при уровне значимости 0,95, так как расчетный критерий Фишера для всех уравнений меньше табличного.

Для определения степени влияния факторов Х1 и Х2 на параметры Y1 и Y2 проведен дисперсионный анализ, который показал, что большее влияние на уровень и колебание натяжения оказывает диаметр эффекта.

Сравнительный анализ полученных тензограмм показал, что минимальный разброс натяжения при сновании данной пряжи обеспечивает двухзонный натяжной прибор НС-1П.

Установлено, что с ростом линейной плотности от 105 до 205 текс прибор НС-1П дает приращение уровня натяжения лишь на 23%, в то время как фарфоровый шайбовый - на 37 %, однозонный НС-1П на 53 %.

При формировании сновальных валов, включающих в себя фасонные и "гладкие" нити, необходима индивидуальная настройка натяжного прибора традиционным методом /11/.

**2.9 Сопутствующая патология при полной утрате зубов у лиц пожилого и старческого возраста**

Изучены эпидемиологически полная утрата зубов и сопутствующая патология пожилого населения, проживающего в домах престарелых на территории Чувашии. Обследование проводилось путем стоматологического осмотра и заполнения статистических карт 784 человек. Результаты анализа показали высокий процент полной утраты зубов, усугубляющейся общей патологией организма. Это характеризует осмотренную категорию населения как группу повышенного стоматологического риска и требует пересмотра всей системы стоматологического обслуживания их.

У пожилых людей уровень заболеваемости в два раза, а в старческом возрасте в шесть раз выше в сравнении с уровнем заболеваемости лиц более молодых возрастов.

Основными заболеваниями лиц пожилого и старческого возраста являются болезни органов кровообращения, нервной системы и органов чувств, органов дыхания, органов пищеварения, костей и органов движения, новообразования и травмы.

Цель исследования – разработка и получение информации о сопутствующих заболеваниях, эффективности зубопротезирования и нуждаемости в ортопедическом лечении лиц пожилого и старческого возраста с полной потерей зубов.

Всего было обследовано 784 человека в возрасте от 45 до 90 лет. Соотношение женщин и мужчин 2,8:1.

Оценка статистической связи с помощью коэффициента корреляции рангов Пирсона позволила установить взаимное влияние отсутствия зубов на сопутствующую заболеваемость с уровнем надежности р=0,0005. Пожилые пациенты с полной потерей зубов страдают болезнями, свойственными старости, а именно, атеросклерозом сосудов головного мозга и гипертонической болезнью.

Дисперсионный анализ показал, что в изучаемых условиях определяющую роль играет специфика болезни. Роль нозологических форм в различных возрастных периодах колеблется в пределах 52—60 %. Наибольшее статистически достоверное влияние на отсутствие зубов оказывают болезни органов пищеварения и сахарный диабет.

В целом группа больных в возрасте 75-89 лет характеризовалась большим числом патологических заболеваний.

В этом исследовании было проведено сравнительное изучение частоты распространения сопутствующей патологии среди пациентов с полной утратой зубов пожилого и старческого возраста, проживающих в домах престарелых. Выявлен высокий процент отсутствия зубов среди лиц этой возрастной категории. У пациентов с полной адентией наблюдается характерная для этого возраста сопутствующая патология. Наиболее часто среди обследованных лиц встречались атеросклероз и гипертония. Статистически достоверно влияние на состояние полости рта таких заболеваний, как болезни желудочно-кишечного тракта и сахарный диабет, доля остальных нозоологических форм оказалась в пределах 52—60 %. Применение дисперсионного анализа не подтвердили значимой роли пола и местожительства на показатели состояния полости рта.

Таким образом, в заключении следует отметить, что анализ распределения сопутствующих заболеваний у лиц с полным отсутствием зубов в пожилом и старческом возрасте показал, что эта категория граждан относится к особой группе населения, которая должна получать адекватную стоматологическую помощь в рамках существующих стоматологических систем /12/.

**3 Дисперсионный анализ в контексте статистических методов**

Статистические методы анализа – это методология измерения результатов деятельности человека, то есть перевода качественных характеристик в количественные.

Основные этапы при проведении статистического анализа:

- содержательный анализ исследуемого объекта, системы или процесса. На этом этапе определяется набор входных и выходных параметров (X1 ,..., Xp; Y1 ,..., Yq);

- составление плана сбора исходных данных - значений входных переменных (X1,...,Xp), числа наблюдений n. Этот этап выполняется при активном планировании эксперимента.

- получение исходных данных и ввод их в компьютер. На этом этапе формируются массивы чисел (x1i ,..., xpi ; y1i ,..., yqi), i=1,..., n, где n - объем выборки.

- первичная статистическая обработка данных. На данном этапе формируется статистическое описание рассматриваемых параметров:

а) построение и анализ статистических зависимостей;

б) корреляционный анализ предназначен для оценивания значимости влияния факторов (X1,...,Xp) на отклик Y;

в) дисперсионный анализ используется для оценивания влияния на отклик Y неколичественных факторов (X1,...,Xp) с целью выбора среди них наиболее важных;

г) регрессионный анализ предназначен для определения аналитической зависимости отклика Y от количественных факторов X;

- интерпретация результатов в терминах поставленной задачи /13/.

В таблице 3.1 приведены статистические методы, с помощью которых решаются аналитические задачи. В соответствующих ячейках таблицы находятся частоты применения статистических методов:

- метка «-» - метод не применяется;

- метка «+» - метод применяется;

- метка «++» - метод широко применяется;

- метка «+++» - применение метода представляет особый интерес /14/.

Дисперсионный анализ подобно t-критерию Стьюдента, позволяет оценить различия между выборочными средними; однако, в отличие от t-критерия, в нем нет ограничений на количество сравниваемых средних. Таким образом, вместо того, чтобы поставить вопрос о различии двух выборочных средних, можно оценить, различаются ли два, три четыре, пять или k средних.

Дисперсионный анализ позволяет иметь дело с двумя или более независимыми переменными (признаками, факторами) одновременно, оценивая не только эффект каждой из них по отдельности, но и эффекты взаимодействия между ними /15/.

Таблица 3.1 – Применение статистических методов при решении аналитических задач

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Аналитические задачи, возникающие в сфере бизнеса, финансов и управления | Методы описательной статистики | Методы поверки статисти-ческих гипотез | Методы регресси-онного анализа | Методы дисперси-онного анализа | Методы анализа категории-альных данных | Методы много-мерного анализа | Методы дискриминантного анализа | Методыкластер-ногоанализа | Методы анализавыжива-емости | Методы анализаи прогнозавременных рядов |
| Задачи горизонталь-ного (временного) анализа | ++ | + | - | + | + | - | - | - | - | - |
| Задачи вертикального (структурного) анализа | ++ | - | - | + | ++ | ++ | + | + | - | - |
| Задачи трендового анализа и прогноза | ++ | - | +++ | ++ | - | - | - | - | ++ | +++ |
| Задачи анализа относительных показателей | ++ | + | + | - | + | +++ | ++ | ++ | - | ++ |
| Задачи сравнительного (пространствен-ного) анализа | ++ | - | + | + | ++ | +++ | ++ | ++ | - | + |
| Задачи факторного анализа | + | + | ++ | - | ++ | +++ | + | ++ | - | + |

К большинству сложных систем применим принцип Парето, согласно которому 20 % факторов определяют свойства системы на 80 %. Поэтому первоочередной задачей исследователя имитационной модели является отсеивание несущественных факторов, позволяющее уменьшить размерность задачи оптимизации модели.

Анализ дисперсии оценивает отклонение наблюдений от общего среднего. Затем вариация разбивается на части, каждая из которых имеет свою причину. Остаточная часть вариации, которую не удается связать с условиями эксперимента, считается его случайной ошибкой. Для подтверждения значимости используется специальный тест - F-статистика.

Дисперсионный анализ определяет, есть ли эффект. Регрессионный анализ позволяет прогнозировать отклик (значение целевой функции) в некоторой точке пространства параметров. Непосредственной задачей регрессионного анализа является оценка коэффициентов регрессии /16/.

Слишком большая размерность выборок затрудняет проведение статистических анализов, поэтому имеет смысл уменьшить размер выборки.

Применив дисперсионный анализ можно выявить значимость влияния различных факторов на исследуемую переменную. Если влияние фактора окажется несущественным, то этот фактор можно исключить из дальнейшей обработки.

**3.1 Векторные авторегрессии**

Макроэконометристы должны уметь решать четыре логически отличающиеся задачи:

- описание данных;

- макроэкономический прогноз;

- структурный вывод;

- анализ политики.

Описание данных означает описание свойств одного или нескольких временных рядов и сообщение этих свойств широкому кругу экономистов. Макроэкономический прогноз означает предсказание курса экономики, обычно на два-три года или меньше (главным образом потому, что прогнозировать на более длинные горизонты слишком трудно). Структурный вывод означает проверку того, соответствуют ли макроэкономические данные конкретной экономической теории. Макроэконометрический анализ политики происходит по нескольким направлениям: с одной стороны, оценивается влияние на экономику гипотетического изменения инструментов политики (например налоговой ставки или краткосрочной процентной ставки), с другой стороны, оценивается влияние изменения правил политики (например переход к новому режиму монетарной политики). Эмпирический макроэкономический исследовательский проект может включать одну или несколько из этих четырех задач. Каждая задача должна быть решена таким образом, чтобы были учтены корреляции между рядами по времени.

В 1970-х годах эти задачи решались с использованием разнообразных методов, которые, если оценить их с современных позиций, были неадекватны по нескольким причинам. Чтобы описать динамику отдельного ряда, достаточно было просто использовать одномерные модели временных рядов, а чтобы описать совместную динамику двух рядов – спектральный анализ. Однако отсутствовал общепринятый язык, пригодный для систематического описания совместных динамических свойств нескольких временных рядов. Экономические прогнозы делались либо с использованием упрощенных моделей авторегрессии — скользящего среднего (ARMA), либо с использованием популярных в то время больших структурных эконометрических моделей. Структурный вывод основывался либо на малых моделях с одним уравнением, либо на больших моделях, идентификация в которых достигалась за счет плохо обоснованных исключающих ограничений, и которые обычно не включали ожидания. Анализ политики на основе структурных моделей зависел от этих идентифицирующих предположений.

Наконец, рост цен в 1970-е годы рассматривался многими как серьезная неудача больших моделей, которые в то время использовались для выработки политических рекомендаций. То есть это было подходящее время для появления новой макроэконометрической конструкции, которая могла бы решить эти многочисленные проблемы.

В 1980 году была создана такая конструкция – векторные авторегрессии (VAR). На первый взгляд, VAR – не более, чем обобщение одномерной авторегрессии на многомерный случай, и каждое уравнение в VAR – не более, чем обычная регрессия по методу наименьших квадратов одной переменной на запаздывающие значения себя и других переменных в VAR. Но этот вроде бы простой инструмент дал возможность систематически и внутренне согласованно уловить богатую динамику многомерных временных рядов, а статистический инструментарий, который сопутствует VAR, оказался удобным и, что очень важно, его было легко интерпретировать.

Выделяют три различных VAR-модели:

- приведенная форма VAR;

- рекурсивная VAR;

- структурная VAR.

Все три являются динамическими линейными моделями, которые связывают текущие и прошлые значения вектора Yt n-мерного временного ряда. Приведенная форма и рекурсивные VAR – это статистические модели, которые не используют никакие экономические соображения за исключением выбора переменных. Эти VAR используются для описания данных и прогноза. Структурная VAR включает ограничения, полученные из макроэкономической теории, и эта VAR используется для структурного вывода и анализа политики.

Приведенная форма VAR выражает Yt в виде распределенного лага прошлых значений плюс серийно некоррелированный член ошибки, то есть обобщает одномерную авторегрессию на случай векторов. Математически приведенная форма модели VAR – это система n уравнений, которые можно записать в матричной форме следующим образом:

 (17)

где  - это n l вектор констант;

 A1, A2, ..., Ap – это n n матрицы коэффициентов;

 t, - это nl вектор серийно некоррелированных ошибок, о которых предполагается, что они имеют среднее ноль и матрицу ковариаций .

Ошибки t, в (17) – это неожиданная динамика в Yt, остающаяся после учета линейного распределенного лага прошлых значений.

Оценить параметры приведенной формы VAR легко. Каждое из уравнений содержит одни и те же регрессоры (Yt–1,...,Yt–p), и нет взаимных ограничений между уравнениями. Таким образом, эффективная оценка (метод максимального правдоподобия с полной информацией) упрощается до обычного МНК, примененного к каждому из уравнений. Матрицу ковариаций ошибок можно состоятельно оценить выборочной ковариационной матрицей полученных из МНК остатков.

Единственная тонкость – определить длину лага p, но это можно сделать, используя информационный критерий, такой как AIC или BIC.

На уровне матричных уравнений рекурсивная и структурная VAR выглядят одинаково. Эти две модели VAR учитывают в явном виде одновременные взаимодействия между элементами Yt, что сводится к добавлению одновременного члена к правой части уравнения (17). Соответственно, рекурсивная и структурная VAR обе представляются в следующем общем виде:

 (18)

где  - вектор констант;

 B0,..., Bp - матрицы;

 t — ошибки.

Наличие в уравнении матрицы B0 означает возможность одновременного взаимодействия между n переменными; то есть B0 позволяет сделать так, чтобы эти переменные, относящиеся к одному моменту времени, определялись совместно.

Рекурсивную VAR можно оценить двумя способами. Рекурсивная структура дает набор рекурсивных уравнений, которые можно оценить с помощью МНК. Эквивалентный способ оценивания заключается в том, что уравнения приведенной формы (17), рассматриваемые как система, умножаются слева на нижнюю треугольную матрицу.

Метод оценивания структурной VAR зависит от того, как именно идентифицирована B0. Подход с частичной информацией влечет использование методов оценивания для отдельного уравнения, таких как двухшаговый метод наименьших квадратов. Подход с полной информацией влечет использование методов оценивания для нескольких уравнений, таких как трехшаговый метод наименьших квадратов.

Необходимо помнить о множественности различных типов VAR. Приведенная форма VAR единственна. Данному порядку переменных в Yt соответствует единственная рекурсивная VAR, но всего имеется n! таких порядков, т.е. n! различных рекурсивных VAR. Количество структурных VAR – то есть наборов предположений, которые идентифицируют одновременные взаимосвязи между переменными, - ограничено только изобретательностью исследователя.

Поскольку матрицы оцененных коэффициентов VAR затруднительно интерпретировать непосредственно, результаты оценивания VAR обычно представляют некоторыми функциями этих матриц. К таким статистикам разложения ошибки прогноза.

Разложения дисперсии ошибки прогноза вычисляются в основном для рекурсивных или структурных систем. Такое разложение дисперсии показывает, насколько ошибка в j-м уравнении важна для объяснения неожиданных изменений i-й переменной. Когда ошибки VAR некоррелированы по уравнениям, дисперсию ошибки прогноза на h периодов вперед можно записать как сумму компонентов, являющихся результатом каждой из этих ошибок /17/.

**3.2 Факторный анализ**

В современной статистике под факторным анализом понимают совокупность методов, которые на основе реально существующих связей признаков (или объектов) позволяют выявлять латентные обобщающие характеристики организационной структуры и механизма развития изучаемых явлений и процессов.

Понятие латентности в определении ключевое. Оно означает неявность характеристик, раскрываемых при помощи методов факторного анализа. Вначале имеется дело с набором элементарных признаков Xj, их взаимодействие предполагает наличие определенных причин, особенных условий, т.е. существование некоторых скрытых факторов. Последние устанавливаются в результате обобщения элементарных признаков и выступают как интегрированные характеристики, или признаки, но более высокого уровня. Естественно, что коррелировать могут не только тривиальные признаки Xj, но и сами наблюдаемые объекты Ni поэтому поиск латентных факторов теоретически возможен как по признаковым, так и по объектным данным.

Если объекты характеризуются достаточно большим числом элементарных признаков (m > 3), то логично и другое предположение - о существовании плотных скоплений точек (признаков) в пространстве n объектов. При этом новые оси обобщают уже не признаки Xj, а объекты ni, соответственно и латентные факторы Fr будут распознаны по составу наблюдаемых объектов:

Fr = c1n1 + c2n2 + ... + cNnN,

где ci - вес объекта ni в факторе Fr.

В зависимости от того, какой из рассмотренных выше тип корреляционной связи - элементарных признаков или наблюдаемых объектов - исследуется в факторном анализе, различают R и Q - технические приемы обработки данных.

Название R-техники носит объемный анализ данных по m признакам, в результате него получают r линейных комбинаций (групп) признаков: Fr=f(Xj), (r=1..m). Анализ по данным о близости (связи) n наблюдаемых объектов называется Q-техникой и позволяет определять r линейных комбинаций (групп) объектов: F=f(ni), (i = l .. N).

В настоящее время на практике более 90% задач решается при помощи R-техники.

Набор методов факторного анализа в настоящее время достаточно велик, насчитывает десятки различных подходов и приемов обработки данных. Чтобы в исследованиях ориентироваться на правильный выбор методов, необходимо представлять их особенности. Разделим все методы факторного анализа на несколько классификационных групп:

- Метод главных компонент. Строго говоря, его не относят к факторному анализу, хотя он имеет с ним много общего. Специфическим является, во-первых, то, что в ходе вычислительных процедур одновременно получают все главные компоненты и их число первоначально равно числу элементарных признаков. Во-вторых, постулируется возможность полного разложения дисперсии элементарных признаков, другими словами, ее полное объяснение через латентные факторы (обобщенные признаки).

- Методы факторного анализа. Дисперсия элементарных признаков здесь объясняется не в полном объеме, признается, что часть дисперсии остается нераспознанной как характерность. Факторы обычно выделяются последовательно: первый, объясняющий наибольшую долю вариации элементарных признаков, затем второй, объясняющий меньшую, вторую после первого латентного фактора часть дисперсии, третий и т.д. Процесс выделения факторов может быть прерван на любом шаге, если принято решение о достаточности доли объясненной дисперсии элементарных признаков или с учетом интерпретируемости латентных факторов.

Методы факторного анализа целесообразно разделить дополнительно на два класса: упрощенные и современные аппроксимирующие методы.

Простые методы факторного анализа в основном связаны с начальными теоретическими разработками. Они имеют ограниченные возможности в выделении латентных факторов и аппроксимации факторных решений. К ним относятся:

- однофакторная модель. Она позволяет выделить только один генеральный латентный и один характерный факторы. Для возможно существующих других латентных факторов делается предположение об их незначимости;

- бифакторная модель. Допускает влияние на вариацию элементарных признаков не одного, а нескольких латентных факторов (обычно двух) и одного характерного фактора;

- центроидный метод. В нем корреляции между переменными рассматриваются как пучок векторов, а латентный фактор геометрически представляется как уравновешивающий вектор, проходящий через центр этого пучка. : Метод позволяет выделять несколько латентных и характерные факторы, впервые появляется возможность соотносить факторное решение с исходными данными, т.е. в простейшем виде решать задачу аппроксимации.

Современные аппроксимирующие методы часто предполагают, что первое, приближенное решение уже найдено каким либо из способов, последующими шагами это решение оптимизируется. Методы отличаются сложностью вычислений. К этим методам относятся:

- групповой метод. Решение базируется на предварительно отобранных каким-либо образом группах элементарных признаков;

- метод главных факторов. Наиболее близок методу главных компонент, отличие заключается в предположении о существовании характерностей;

- метод максимального правдоподобия, минимальных остатков, а-факторного анализа канонического факторного анализа, все оптимизирующие.

Эти методы позволяют последовательно улучшить предварительно найденные решения на основе использования статистических приемов оценивания случайной величины или статистических критериев, предполагают большой объем трудоемких вычислений. Наиболее перспективным и удобным для работы в этой группе признается метод максимального правдоподобия.

Основной задачей, которую решают разнообразными методами факторного анализа, включая и метод главных компонент, является сжатие информации, переход от множества значений по m элементарным признакам с объемом информации n х m к ограниченному множеству элементов матрицы факторного отображения (m х r) или матрицы значений латентных факторов для каждого наблюдаемого объекта размерностью n х r, причем обычно r < m.

Методы факторного анализа позволяют также визуализировать структуру изучаемых явлений и процессов, а это значит определять их состояние и прогнозировать развитие. Наконец, данные факторного анализа дают основания для идентификации объекта, т.е. решения задачи распознавания образа.

Методы факторного анализа обладают свойствами, весьма привлекательными для их использования в составе других статистических методов, наиболее часто в корреляционно-регрессионном анализе, кластерном анализе, многомерном шкалировании и др. /18/.

**3.3 Парная регрессия. Вероятностная природа регрессионных моделей.**

Если рассмотреть задачу анализа расходов на питание в группах с одинаковыми доходами, например в $10.000(x), то это детерминированная величина. А вот Y - доля этих денег, затрачиваемая на питание - случайна и может меняться от года к году. Поэтому для каждого i-го индивида:

где εi - случайная ошибка;

 α и β - константы (теоретически), хотя могут меняться от модели к модели.

Предпосылки для парной регрессии:

- X и Y связаны линейно;

- Х - неслучайная переменная с фиксированными значениями;

- ε - ошибки нормально распределены N(0,σ2);

- ;

- .

На рисунке 3.1 представлена модель парной регрессии.

Рисунок 3.1 – Модель парной регрессии

Эти предпосылки описывают классическую линейную регрессионную модель.

Если ошибка имеет ненулевое среднее, исходная модель будет эквивалентна новой модели и другим свободным членом, но с нулевым средним для ошибки.

Если выполняются предпосылки, то МНК оценки и являются эффективными линейными несмещенными оценками

Если обозначить:

то что математическое ожидание и дисперсии коэффициентов и будут следующие:

Ковариация коэффициентов:

Если то и распределены тоже нормально:

Отсюда следует, что:

- Вариация β полностью определяется вариацией ε;

- Чем выше дисперсия X - тем лучше оценка β.

Полная дисперсия определяется по формуле:

Дисперсия отклонений в таком виде - несмещенная оценка и называется стандартной ошибкой регрессии. N-2 - может быть интерпретировано как число степеней свободы.

Анализ отклонений от линии регрессии может представить полезную меру того, насколько оцененная регрессия отражает реальные данные. Хорошая регрессия та, которая объясняет значительную долю дисперсии Y и наоборот плохая регрессия не отслеживает большую часть колебаний исходных данных. Интуитивно ясно, что всякая дополнительная информация позволит улучшить модель, то есть уменьшить необъясненную долю вариации Y. Для анализа регрессионной модели проводят разложение дисперсии на составляющие, определяют коэффициент детерминации R2.

Отношение двух дисперсий распределено по F-распределению, т. е. если проверить на статистическую значимость отличия дисперсии модели от дисперсии остатков, можно сделать вывод о значимости R2.

Проверка гипотезы о равенстве дисперсий этих двух выборок:

Если гипотеза Н0 (о равенстве дисперсий нескольких выборок) верна, t имеет F-распределение с (m1,m2)=(n1-1,n2-1) степенями свободы.

Посчитав F – отношение как отношение двух дисперсий и сравнив его с табличным значением, можно сделать вывод о статистической значимости R2 /2/, /19/.

**Заключение**

Современные приложения дисперсионного анализа охватывают широкий круг задач экономики, биологии и техники и трактуются обычно в терминах статистической теории выявления систематических различий между результатами непосредственных измерений, выполненных при тех или иных меняющихся условиях.

Благодаря автоматизации дисперсионного анализа исследователь может проводить различные статистические исследования с применение ЭВМ, затрачивая при этом меньше времени и усилий на расчеты данных. В настоящее время существует множество пакетов прикладных программ, в которых реализован аппарат дисперсионного анализа. Наиболее распространенными являются такие программные продукты как:

- MS Excel;

- Statistica;

- Stadia;

- SPSS.

В современных статистических программных продуктах реализованы большинство статистических методов. С развитием алгоритмических языков программирования стало возможным создавать дополнительные блоки по обработке статистических данных.

Дисперсионный анализ является мощным современным статистическим методом обработки и анализа экспериментальных данных в психологии, биологии, медицине и других науках. Он очень тесно связан с конкретной методологией планирования и проведения экспериментальных исследований.

Дисперсионный анализ применяется во всех областях научных исследований, где необходимо проанализировать влияние различных факторов на исследуемую переменную.

**Список литературы**

1 Кремер Н.Ш. Теория вероятности и математическая статистика. М.: Юнити – Дана, 2002.-343с.

2 Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высшая школа, 2003.-523с.

3 www.sutd.ru

4 www.conf.mitme.ru

5 www.pedklin.ru

6 www.webcenter.ru

7 www.infections.ru

8 www.encycl.yandex.ru

9 www.infosport.ru

10 www.medtrust.ru

11 www.flax.net.ru

12 www.jdc.org.il

13 www.big.spb.ru

14 www.bizcom.ru

15 Гусев А.Н. Дисперсионный анализ в экспериментальной психологии. – М.: Учебно-методический коллектор «Психология», 2000.-136с.

16 www.gpss.ru

17 www.econometrics.exponenta.ru

18 www.optimizer.by.ru

19 www2.econ.msu.ru